DOCUMENT RESUME

ED 312 296 TM 014 068

AUTHOR Benton, Roberta
TITLE Planned Comparisons as Better Alternatives to ANOVA Omnibus Tests.
PUB DATE Nov 89
NOTE 19p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Little Rock, AR, November 8-10, 1989).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Analysis of Variance; Comparative Analysis; *Research Design; Statistical Analysis
IDENTIFIERS A Priori Tests; Coding; *Omnibus Tests; Orthogonal Comparison; *Planned Comparisons; Post Hoc Tests; Power (Statistics); Type II Errors

ABSTRACT

Analyses of data are presented to illustrate the advantages of using a priori or planned comparisons rather than omnibus analysis of variance (ANOVA) tests followed by post hoc or posteriori testing. The two types of planned comparisons considered are planned orthogonal non-trend coding contrasts and orthogonal polynomial or trend contrast coding. A hypothetical one-way design was devised to illustrate the superiority of planned comparisons. The data set used one dependent variable with five levels. Ten subjects were assigned evenly, two to each of five groups. The dependent variable consisted of hypothetical scores on a measure of performance. The examples demonstrate that planned comparisons using orthogonal coding may be preferable to classical ANOVA followed by post hoc testing. These procedures: (1) provide greater statistical power against Type II error; (2) help isolate specific sources of variance in the dependent variable; (3) help determine forms or patterns of relationships; and (4) make the researcher think about relationships when designing proposed studies. Researchers should consider using planned tests more often when factors have more than two levels. Five data tables and two appendices on non-trend comparisons and trend coding contrasts supplement the text. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED312296

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

[✓] This document has been reproduced as received from the person or organization originating it.

[ ] Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ROBERTA BENTON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Planned Comparisons as Better Alternatives to ANOVA Omnibus Tests

Roberta Benton

Louisiana State University Medical Center

Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock, AR, November 9, 1989.

TM014068

BEST COPY AVAILABLE

Analysis of variance methods (ANOVA) are among the most commonly applied methods in social science research (Elmore & Woehlke, 1988; Gaither & Glorfeld, 1985; Goodwin & Goodwin, 1985; Thompson, 1986), yet there are occasions when "ANOVA through regression" using planned comparisons can be superior to classical ANOVA followed by post hoc testing. This paper presents several analyses of example data for the purpose of illustrating the alternative of using a priori or planned comparisons in preference to omnibus ANOVA tests followed by post hoc or a posteriori testing.

Classical ANOVA methods traditionally test all possible omnibus effects or relationships among the variables. If a statistically significant result is obtained for a way involving more than two levels, then the researcher using classical ANOVA methods must seek the source of the effect. Post hoc or a posteriori tests are applied to investigate such omnibus results. Various methodologists have referred to this post hoc testing practice as "fishing" (Minium & Clarke, 1982, p. 321), "milking" (Keppel, 1982, p. 150) or "data snooping" (Kirk, 1968, p. 73; Kirk, 1984, p. 360; Pedhazur, 1982, p. 305).

Classical ANOVA With Post Hoc Unplanned Testing

A hypothetical "one way" (Kerlinger, 1979, p. 84) design was devised for the purposes of this paper to illustrate the superiority of planned comparisons over omnibus ANOVA testing. The data set employed one

independent variable with five levels. Ten subjects were assigned evenly, two to each of the five groups, thus creating a balanced design. The dependent variable consisted of hypothetical scores on a measure of performance.

Table 1 presents the results of a classical omnibus ANOVA. The procedure indicated no statistically significant differences among the means of the five groups (F=1.8, df=4/5, p=0.266). The eta-squared effect size, calculated by determining the ratio of the main effect sum of squares (SOS) to the total sum of squares, was 59% ((360/610) x 100).

INSERT TABLE 1 ABOUT HERE

Although post hoc testing is done only when statistically significant results are obtained (Pedhazur, 1982), for illustrative purposes the data were subjected to seven commonly utilized a posteriori methods (LSD, Duncan, SNK, Tukey B, Tukey, LSDMOD, Scheffe). None of the post hoc or unplanned tests detected that any two group means were significantly different at the 0.05 level. Post hoc analyses were performed here to illustrate that even these procedures did not reveal the existing differences between means for the example data.

Planned Comparisons

There are several reasons for utilization of orthogonal comparisons and "ANOVA through regression" instead of

classical ANOVA in which omnibus main effects are tested as a composite rather than being partitioned into separate components. One important advantage of planned comparisons is the protection such a strategy provides against Type II error, i.e., the failure to detect true differences. In comparing a priori or planned versus post hoc or unplanned comparisons Kerlinger and Pedhazur (1973, p. 131) state:

> The tests of significance for a priori, or planned, comparisons are more powerful than those for post hoc comparisons. In other words, it is possible for a specific comparison to be not significant when tested by post hoc methods but significant when tested by a priori methods.

As Thompson (1989, p. 7) reiterates, "Regression approaches to ANOVA using a priori contrasts have more statistical power against Type II error."

Another reason for utilization of planned comparisons is more exact determination of the sources of variance in the dependent variable arising from the predictor variables. The researcher is able to gain better insight into relationships among the variables.

Perhaps the most important reason for employing a priori or focused comparisons is that planned comparisons force the researcher to think about the focus of the study, how the variables may be related, and all the possible outcomes in a given study, because only a limited number of

uncorrelated or orthogonal planned comparisons can be set forth in a given study (Thompson, 1988). As Hays (1981) concurs, planned comparisons partition dependent variable variance into segments which focus on the hypotheses believed to be important in the experiment. The two types of planned comparisons to be considered here are planned orthogonal non-trend contrasts and orthogonal polynomial or trend contrasts.

Orthogonal Non-trend Coding Contrasts

Hicks (1973) defines a contrast as "a linear combination of treatment totals or averages where the sum of the coefficients is zero" and explains that "two contrasts are said to be orthogonal if the products of their corresponding coefficients add to zero" (Hicks, 1973, p. 305).

The first step in the analysis of the hypothetical data using orthogonal non-trend contrasts was to assign orthogonal effect coding to each of the five levels of the independent variable according to the methods suggested by Kerlinger (1973), Thompson (1988), and others. Coding partitions the omnibus sum of squares of the dependent variable into smaller parts using one degree of freedom hypotheses, an important consideration when a way or factor has more than two levels (Thompson, 1985).

Table 2 illustrates the coding scheme. The method utilizes assignment of -1 to each subject in Level 1, +1 to

each subject in Level 2, and 0 to the remaining subjects; this contrast was designated "A1". In the second contrast or "probe" (Thompson, 1989, p. 5), -1 is assigned to each subject who received a non-zero numerical value in contrast #1. A +2 (for the number of the probe) is given to each subject in the third level of the way, and this contrast was labelled "A2". Remaining subjects of the second probe receive the code of 0. The same rules apply to coding for probes #3 (A3) and #4 (A4). The last probe for the way in this coding scheme involves no zeroes, which means that data from each level of the way are considered when testing the hypothesis associated with this constrast. The total number of probes or contrasts is one less than the number of levels in the independent variable. Each orthogonal contrast is assigned one degree of freedom. Each probe explains a unique, uncorrelated, nonoverlapping portion of the variance of the dependent variable (Thompson, 1989).

INSERT TABLE 2 ABOUT HERE

By analyzing the coding presented in Table 2 the researcher can determine the group means that were tested in a given contrast hypothesis. For example, the "A1" contrast tested the equality of the dependent variable mean of level 1 subjects versus the mean of level 2 subjects (Ho: u of level 1=u of level 2). The "A4" hypothesis tested the equality of the mean from a combination of level 1 through level 4 subjects versus the dependent variable mean of level

5 subjects (Ho: [u of level 1 + u of level 2 + u of level 3 + u of level 4] /4 = u of level 5).

Assignment of coding numbers can be varied depending of which groups the researcher wishes to compare based on previous research or theory; however, in this coding scheme all probes must sum to zero and all crossproducts of any two contrasts must sum to zero.

All of the explained effect size demonstrated in the classical ANOVA was isolated solely from the "A4" probe in the "ANOVA through regression" analysis. The resulting F ratio was statistically significant (F=7.2, df=1/5, p<.05). The effect size for the "A4" contrast was 59% ((360/610) x 100), as was the effect size for the omnibus main effect hypothesis in the classical ANOVA. The total explained variance was realized by the comparison of the mean of the level 5 subjects with the dependent variable mean of the remaining subjects in levels 1 through 4. Table 3 presents the results of hand calculation of the keyout utilizing computer generated data from an "ANOVA through regression" using this coding scheme. The SPSS-X commands for computer generation of orthogonal coding are included in Appendix A.

INSERT TABLE 3 ABOUT HERE

Orthogonal Polynomial or Trend Contrast Coding

If intervals of the predictor variable are equally spaced and the sample size is constant for all levels, trend coding can provide information about the general form of the

relationship between the dependent variable and the levels of the independent variable (Winer, 1971). Snedecor (1980) presents a readable account of polynomial response curves or trends and explains that when the relation between the predictor variable and the dependent variable is curved, a complex mathematical expression is required to describe it. He explains the mathematical derivation for the coding which results in linear, quadratic, cubic, quartic, and quintic coefficients which are orthogonal. In each coding sequence, the number of times that the signs change determines the degree of the polynomial. For example: linear (the first degree equation) has only one sign change during the sequence (-5 -3 -1 +1 +3 +5); in the cubic contrast (third degree equation) there are three changes of signs (-5 +7 +4 -4 -7 +5) (Winer, 1971). When represented graphically, each component will have one less bend or curve than the power that each contrast or planned comparison represents. For example, the quadratic contrast, which involves squaring or raising to the second power, will possess one bend or curve in the pattern created by graphing the means.

By utilizing polynomial contrasts the sum of squares for the main effects of a factor can be divided into linear, quadratic, cubic, quartic, or quintic components. Each sum of squares is tested against the error SOS/error df (i.e., mean square) to evaluate the type of pattern that describes

the response curve or trend in the group means across the levels of the way or factor (Snedecor, 1980).

Contrasts were created to test for trends in the means, utilizing the table, Coefficients of Orthogonal Polynomials, from Hicks (1973, p. 324). Similar tables of trend contrasts for various designs are presented by Winer (1971, p. 878) and by Snedecor (1980, p. 491). The hypothetical five level way in the present example was coded using linear, quadratic, cubic, and quartic trend coding. Four contrasts are again required since a five level way has four degrees of freedom and thus requires four contrasts. Table 4 lists the trend coding output from the SPSS-X generated contrasts and the "IF" and "COMPUTE" statements for generating orthogonal trend contrasts are included in Appendix B.

INSERT TABLE 4 ABOUT HERE

The Pearson correlation coefficients among the contrast variables comfirms that the contrasts are truly exactly orthogonal. Each probe sums to zero and the sums of the cross products of each contrast with any other contrast is zero, so this result merely confirms that the coding was correctly done. Table 5 reports the hand calculated keyout for trend analysis derived from the computer generated output. Observation reveals that each level of the predictor variable is correlated with the dependent variable in differing degrees. The linear component defines 29.5% of the explained variance; the quadratic component, 21.08%; the

cubic component, 7.38%; and quartic component, 1.05%. The total of all these effects equals the total effect size of the main effect in the classical ANOVA (59%). None of the F statistic values were significant at the 0.05% level.

INSERT TABLE 5 ABOUT HERE

In the trend coding of the hypothetical data set, the best fitting curve appears to contain a combination of linear and quadratic components. The linear sum of squares (180) and the quadratic sum of squares (128.57) combine to account for 308.57 of the 360 main effect total variance.

This result is only to be expected, given the actual pattern of the dependent variable means across the five levels for the hypothetical data reported in Table 2, i.e., 10, 10, 10, 10, and 25. This pattern is largely linear, and would have been perfectly linear if the mean in the fifth level had been 10 rather than 25. The means, when plotted, do create a line with one bend or curve, which occurs between the fourth and fifth levels. Thus, the pattern is partially quadratic as well. The cubic and quartic components (45+6.43=51.43) make up the remaining variation in the criterion variable.

Winer (1971, p. 178) advises caution in the use of polynomials and emphasizes that "the nature of the data and the purpose of the experiment must be considered in evaluation of the results". However, in studies where

various trends from intervention effects are expected, this coding scheme can be quite useful.

Summary

The examples provided in this paper illustrate occasions when planned comparisons utilizing orthogonal coding may be preferable to classical ANOVA followed by post hoc testing. These procedures provide greater statistical power against Type II error, help isolate specific sources of variance in the dependent variable, help determine forms or patterns of relationships, and force the researcher to think about relationships when designing proposed studies. Sowell and Casey (1982, p. 119) support this thesis in stating that "planned comparisons are the most powerful comparison tests available," and none of the post hoc procedures "has the power of planned comparison tests for detecting statistical significance". Therefore, researchers should more frequently consider using planned tests when ways or factors have more than two levels.

References

Elmore, P. B., & Woehlke, P.L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Edwards, A. L. (1979). Multiple regression and the analysis of variance and covariance. San Francisco: W. H. Freeman.

Gaither, N., & Glorfeld, L. (1985). An evaluation of the use of tests of significance in organizational behavior research. Academy of Management Review, 10, 787-793.

Goodwin, L.D., & Goodwin, W.L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

Hays, W. L. (1981). Statistics (3rd ed.). New York: Holt, Rinehart and Winston.

Hicks, C. R. (1973). Fundamental concepts in the design of experiments. New York: Holt, Rinehart and Winston.

Keppel, G. (1982). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice-Hall.

Kerlinger, F. N. (1973). Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

Kerlinger, F. N. (1979). Behavioral research: A conceptual approach. New York: Holt, Rinehart and Winston.

Kirk, R. E. (1968). Experimental design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole.

Kirk, R. E. (1984). Elementary statistics. Monterey, CA: Brooks/Cole.

Minium, W. W., & Clarke, R. B. (1982). Elements of statistical reasoning. New York: John Wiley.

Pedhazur, E. J. (1982). Multiple regression in behavioral research (2nd ed.). New York: Holt, Rinehart and Winston.

Snedecor, G. W., & Cochran, W. G. (1980). Statistical methods (7th ed.). Ames, IA: The Iowa State University Press.

Sowell, E. J., & Casey, R. J. (1982). Research methods in education. Belmont, CA: Wadsworth.

Thompson, B. (1985). Alternate methods for analyzing data from education experiments. Journal of Experimental Education, 54, 50-55.

Thompson, B. (1986). ANOVA versus regression analysis of ATI designs: An empirical investigation. Educational and Psychological Measurement, 46, 917-928.

Thompson, B. (1988). The importance of planned or focused comparisons in OVA research. Measurement and Evaluation in Counseling and Development, 21, 99-101.

Thompson, B. (1989, January). Heuristics for understanding the concepts of interaction, polynomial trend, and the general linear model. Paper presented at the meeting of the Southwest Educational Research Association, Houston, TX. (ERIC Document Reproduction Service No. ED 304 453)

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Table 1

Keyout for Classical ANOVA Analysis

Source	SOS	DF	Mean Square	F	Sig F	Effect Size
Explained	360	4	90.00	1.8	.266	59.00%
Residual	250	5	50.00			
Total	610	9	67.78			

Table 2

Non-trend Coding Contrasts

LEV	DV	A1	A2	A3	A4
1	5	-1.00	-1.00	-1.00	-1.00
1	15	-1.00	-1.00	-1.00	-1.00
2	5	1.00	-1.00	-1.00	-1.00
2	15	1.00	-1.00	-1.00	-1.00
3	5	.00	2.00	-1.00	-1.00
3	15	.00	2.00	-1.00	-1.00
4	5	.00	.00	3.00	-1.00
4	15	.00	.00	3.00	-1.00
5	20	.00	.00	.00	4.00
5	30	.00	.00	.00	4.00

Table 3
"ANOVA through Regression" Using Non-trend Contrasts

Source	SOS	DF	Mean Square	F	Sig F	Effect Size
A1	0	1	0.00	0.0		0.00%
A2	0	1	0.00	0.0		0.00%
A3	0	1	0.00	0.0		0.00%
A4	360	1	360.00	7.2	.005	59.00%
Residual	250	5	50.00			
Total	610	9	67.78			

Table 4
Trend Coding Contrasts

LEV	DV	LIN	QUAD	CUB	QUART
1	5	-2.00	2.00	-1.00	1.00
1	15	-2.00	2.00	-1.00	1.00
2	5	-1.00	-1.00	2.00	-4.00
2	15	-1.00	-1.00	2.00	-4.00
3	5	.00	-2.00	.00	6.00
3	15	.00	-2.00	.00	6.00
4	5	1.00	-1.00	-2.00	-4.00
4	15	1.00	-1.00	-2.00	-4.00
5	20	2.00	2.00	1.00	1.00
5	30	2.00	2.00	1.00	1.00

Table 5
Keyout from Regression Using Trend Contrasts

Source	SOS	DF	Mean Square	F	Effect Size
LIN	180.000	1	180.000	3.600	29.50%
QUAD	128.571	1	128.571	2.571	21.08%
CUB	45.000	1	45.000	0.900	7.38%
QUART	6.428	1	6.428	0.129	1.05%
Residual	250.000	5	50.000		
Total	610.000	9	67.777		

APPENDIX A:
SPSS-X CONTROL CARDS FOR NON-TREND COMPARISONS

```
TITLE 'A PRIORI VS OMNIBUS WITH POST HOC'
FILE HANDLE RB/NAME='APRIORI.DAT'
DATA LIST FILE=RB/LEV 1 DV 2-3
LIST VARIABLES=LEV TO DV/CASES 900
REGRESSION VARIABLES=LEV DV/DESCRIPTIVES=ALL/DEPENDENT=DV/
  ENTER LEV
SUBTITLE 'CLASSICAL ANOVA'
ANOVA DV BY LEV (1,5)
LIST VARIABLES=ALL/CASES=900
SUBTITLE 'ANOVA THROUGH REGRESSION'
COMPUTE A1=0
IF (LEV EQ 1) A1=-1
IF (LEV EQ 2) A1=1
IF (A1 NE 0) A2=-1
IF (LEV EQ 3) A2=2
IF (LEV GE 4) A2=0
IF (A2 NE 0) A3=-1
IF (LEV EQ 4) A3=3
IF (LEV GE 5) A3=0
IF (A3 NE 0) A4=-1
IF (LEV EQ 5) A4=4
LIST VARIABLES=ALL/CASES=900
SUBTITLE 'DEMONSTRATE CONTRASTS ARE PERFECTLY UNCORRELATED'
CORRELATIONS VARIABLES=DV A1 A2 A3 A4
SUBTITLE 'SHOW ENTRY ONE AT A TIME NOT CHANGE RESULTS'
REGRESSION VARIABLES=DV A1 TO
A4/DESCRIPTIVES=DEFAULT/DEPENDENT=DV/
  ENTER A1/ENTER A2/ENTER A3/ENTER A4
SUBTITLE 'POST HOC ANALYSES'
ONEWAY DV BY LEV(1,5)/RANGES=LSD/RANGES=DUNCAN/
  RANGES=SNK/RANGES=TUKEYB/
  RANGES=TUKEY/RANGES=LSDMOD/RANGES=SCHEFFE/HARMONIC=ALL
```

APPENDIX B:
IF AND COMPUTE STATEMENTS FOR TREND CODING CONTRASTS

```
COMPUTE LIN=0
IF (LEV EQ 1) LIN=-2
IF LEV EQ 2) LIN=-1
IF (LEV EQ 4) LIN=1
IF (LEV EQ 5) LIN=2
COMPUTE QUAD=2
IF (LEV EQ 2) QUAD=-1
IF (LEV EQ 3) QUAD=-2
IF (LEV EQ 4) QUAD=-1
COMPUTE CUB=0
IF (LEV EQ 1) CUB=-1
IF (LEV EQ 2) CUB=2
IF (LEV EQ 4) CUB=-2
IF (LEV EQ 5) CUB=1
COMPUTE QUART=1
IF (LEV EQ 2) QUART=-4
IF (LEV EQ 3) QUART=6
IF (LEV EQ 4) QUART=-4
```